BBC RD 1993/2

Research Department Report

DIGITAL SOUND:

Subjective tests on commentary-quality codecs

N.H.C. Gilchrist, B.Sc., C.Eng., M.I.E.E. *and* **A.J. Oxenham**

Research Department, Engineering Division
THE BRITISH BROADCASTING CORPORATION

DIGITAL SOUND: SUBJECTIVE TESTS ON COMMENTARY-QUALITY CODECS

N.H.C. Gilchrist, B.Sc., C.Eng., M.I.E.E. and A.J. Oxenham

Summary

Subjective tests have been carried out on five commentary-quality audio codecs on behalf of CCIR Task Group 10/2. All codecs operated at a bit rate of 60 kbit/s per mono signal, and were tested in both mono and stereo configurations.

Significant differences in performance were found. All codecs introduced some impairment on the most critical programme items.

Issued under the Authority of

Ian Childs

Head of Research Department

Research Department, Engineering Division
BRITISH BROADCASTING CORPORATION

(S-18)

1993

© British Broadcasting Corporation

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without prior permission.

DIGITAL SOUND: SUBJECTIVE TESTS ON COMMENTARY-QUALITY CODECS

N.H.C. Gilchrist, B.Sc., C.Eng., M.I.E.E. and A.J. Oxenham

1. Introduction 1
2. Brief Details of the Codecs 1
3. Selection of Critical Programme Material 1
4. Preparation of the Recordings for Subjective Tests 1
5. Description of the Tests 3
6. Discussion of Results 3
7. Conclusions 4
8. Acknowledgements 4
9. References 4

Appendix 6

© BBC 2003. All rights reserved. Except as provided below, no part of this document may be reproduced in any material form (including photocopying or storing it in any medium by electronic means) without the prior written permission of BBC Research & Development except in accordance with the provisions of the (UK) Copyright, Designs and Patents Act 1988.

The BBC grants permission to individuals and organisations to make copies of the entire document (including this copyright notice) for their own internal use. No copies of this document may be published, distributed or made available to third parties whether by paper, electronic or other means without the BBC's prior written permission. Where necessary, third parties should be directed to the relevant page on BBC's website at http://www.bbc.co.uk/rd/pubs/ for a copy of this document.

DIGITAL SOUND: SUBJECTIVE TESTS ON COMMENTARY-QUALITY CODECS

N.H.C. Gilchrist, B.Sc., C.Eng., M.I.E.E. and A.J. Oxenham*

1. INTRODUCTION

Techniques for reducing the bit-rate requirements of digital audio signals have been a concern of CCIR** Study Group 10 for many years, and in 1990 the Study Group formed Task Group 10/2† to study low bit rate audio coding and prepare a draft Recommendation on this subject. The terms of reference for the Task Group are contained in CCIR Decision 98[1].

An essential part of the Task Group's work is the conduct of listening tests to evaluate the performance of different coding systems. There is also a strong interest in the development of measurement methods for assessing audio quality, but as yet these cannot be relied upon to the extent that they could replace listening tests.

This Report describes the listening tests conducted by BBC Research Department to assess the basic audio quality of codecs intended for use with speech contribution signals for commentary and NCA (News and Current Affairs) and for medium-quality mono and stereo programme distribution, including music. The identity of the individual codecs is not revealed, and they are designated A, B, C, D and E in the Report.

2. BRIEF DETAILS OF THE CODECS

All of the codecs under test exchanged digital audio signals with the signal sources and the recording equipment in AES/EBU (CCIR Recommendation 647) form[2]. Four of the five codecs operated at a sampling frequency of 48 kHz; codec D operated at a sampling frequency of 32 kHz.

Task Group 10/2 requires all commentary codecs to code the audio data with a bit rate of 60 kbit/s per mono channel. This allows an extra 4 kbit/s per mono channel to be used for error correction and ancillary data, while the overall bit rate remains within the 64 kbit/s provided by an ISDN B channel.

All of the stereo codecs (except for codec A) incorporate algorithms which exploit correlation between the two components of a stereo signal. Theoretically, the quality of a stereo signal should be somewhat better than a mono signal passed through only one channel of a codec. Because of this, the tests were divided into two categories, mono and stereo. The same stereo programme material was used as the source for both sets of tests, but a mono mix was derived for the mono tests.

3. SELECTION OF CRITICAL PROGRAMME MATERIAL

The arrangement shown in the block schematic diagram of Fig. 1 *(overleaf)* was used for the selection of critical material. Three listeners, from the Swedish Broadcasting Corporation, the CCETT (Centre Commune d'Études de Télédiffusion et Télécommunications) and the BBC, listened to a wide range of programme material processed by the codecs. This included items which had been found critical in earlier listening tests by the three organisations, material previously used in Eureka 147 tests on low bit-rate codecs[3], material collected for the ISO Moving Picture Experts' Group (MPEG)[4, 5] and material recommended for consideration by the various codec proponents and members of Task Group 10/2. Eight items were selected for commentary codec testing, and these are listed in Table 1 *(overleaf)*.

During the selection of critical material, the listeners observed that it was not always the most 'clean' (i.e. noise-free and distortion-free) material which proved to be the most critical. Sometimes material which was somewhat imperfect was the most effective; the material selected included some imperfect items. It was evident, too, when the same material was available both as a Compact Disc and as a DAT recording, that after several generations of digital copying (in this case, an ISO MPEG recording), the DAT recording seemed to be a more critical source than the original CD.

4. PREPARATION OF THE RECORDINGS FOR SUBJECTIVE TESTS

Once the critical test material had been chosen, recordings of the unprocessed source material were

---

* Final-year undergraduate student, Music Department, University of Surrey, at the time of these tests.

** Comité Consultatif Internationale des Radiocommunications, the International Radio Consultative Committee (part of the International Telecommunication Union). On 1st March 1993, the CCIR was renamed 'Radiocommunication Bureau'.

† The Task Group was formed in association with the CMTT, and was known initially as Joint Interim Working Party 10/CMTT/1. The change in title resulted from a CCIR reorganisation.

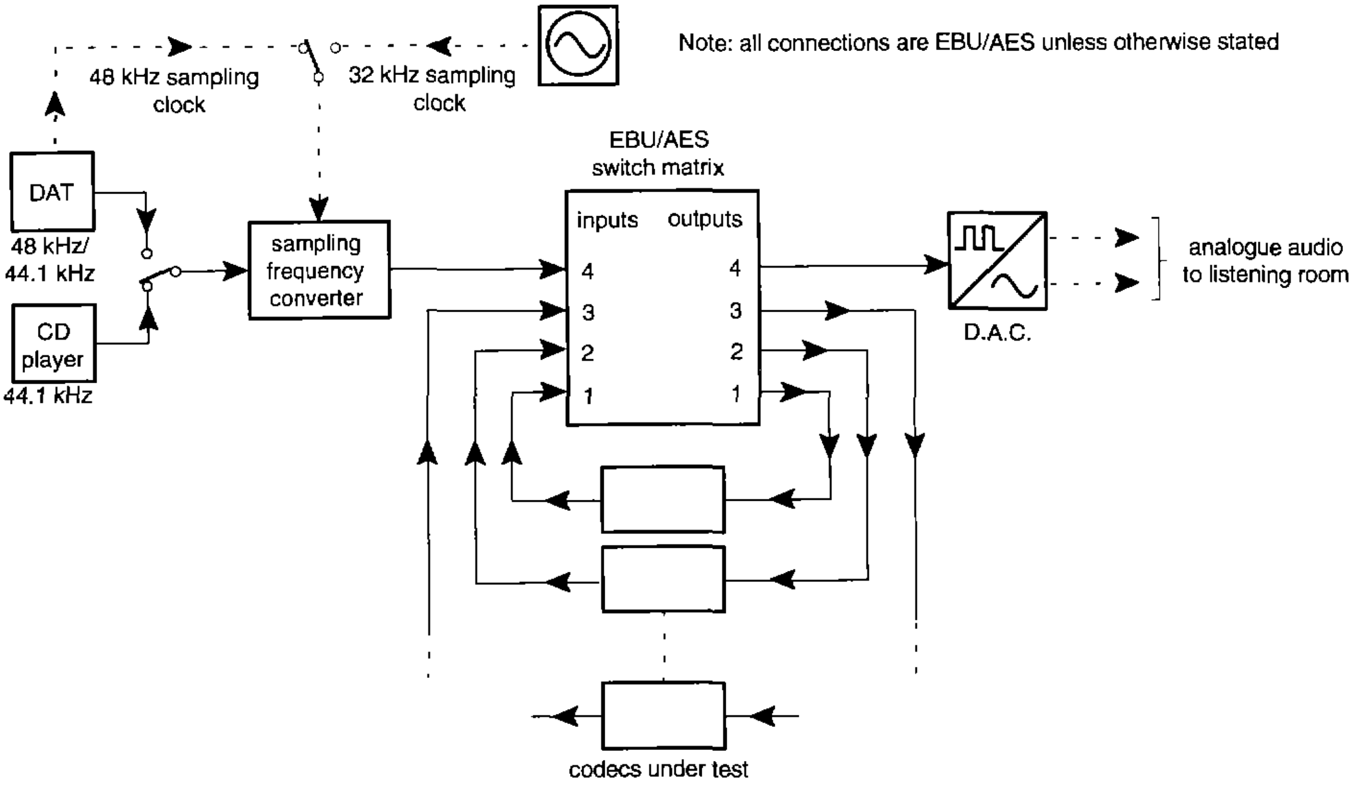


Fig. 1 - Arrangement for testing codecs with recorded programme material

assembled on digital audio tape (DAT) at 48 kHz sampling rate with the aid of a digital audio editor[6]. For the mono tests, mono mixes (a summation of left and right channels of the stereo material, each attenuated by 6 dB) were required. As a precaution against the exploitation of channel-to-channel correlation by a mono commentary codec providing two channels in 120 kbit/s, the mono mix was sent in only one set of the sub-frames of EBU/AES interface; a mono mix of 'pop' music at a high level occupied the other sub-frames. The various mono mixes were made in a BBC six-channel HDTV digital audio matrix[7], which was also used to place the mono mixes, as required, in the left and right sub-frames of the EBU/AES interface. Fig. 2 shows the interconnection of the matrix with DAT recorders and a CD player for preparing the mono recordings.

Recordings of signals processed by a codec were made by simply connecting the codec between two DAT recorders. The codec operating at a sampling frequency of 32 kHz was connected via

Table 1: Critical material for commentary codecs.

Item no.	Description		Duration (seconds)
1.	Åsa Jinder: Stilla Skymning, track 5.	(Eagle Records. CD No. ECD 015)	25
2.	Solo harpischord *arpeggio*, track 40.1.	(EBU SQAM CD. No. 422-204-2)	9
3.	Castanets, track 27.	(EBU SQAM CD. No. 422-204-2)	16
4.	German male speech, track 54.	(EBU SQAM CD. No. 422-204-2)	17
5.	Bass clarinet *arpeggio*, track 17.	(EBU SQAM CD. No. 422-204-2)	10
6.	Ornette Coleman: 'In All Languages', track 7.	('The Original Quartet and Prime Time', on Dreams CD. No. 008)	19 19
7.	K.B. Enchill: solo base guitar.	(Swedish National Radio Recording)	20
8.	Suzanne Vega, track 1.	(A&M CD. No. 395 136-2)	20

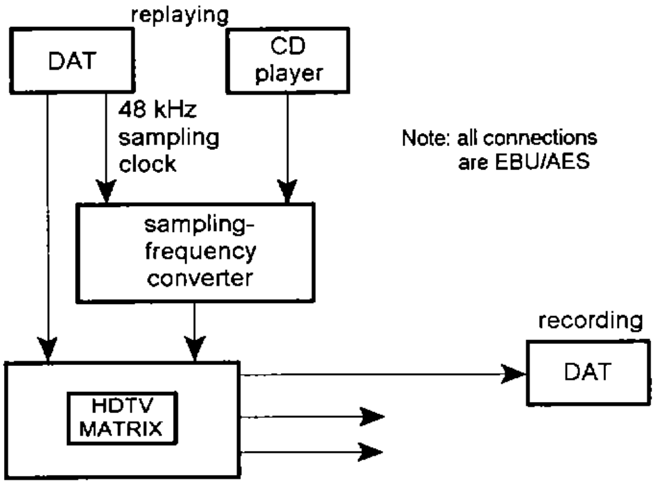


Fig. 2 - Arrangement for preparing mono material.

sampling-frequency converters, and all DAT recordings were made and replayed at 48 kHz sampling frequency.

A specially-recorded introduction and some test-number announcements were also transferred to the editor. The stored material was then compiled into test sequences and transferred digitally back to DAT format.

5. DESCRIPTION OF THE TESTS

Prior to conducting the listening tests, measurements were made on the codecs to determine the bit rate exchanged between coder and decoder, and the delay introduced into the audio signal by each codec[8]. The amplitude/frequency response of each codec was also measured*.

For the listening tests, the sequences of processed and unprocessed critical programme material were assembled using the disk-based editor, and re-recorded on DAT. The DAT recordings were replayed in a listening room where the sound was reproduced using BBC monitoring loudspeakers type LS5/8.

The listening room had a low ambient noise level, representative of good domestic listening conditions, a mid-band reverberation time of about 0.3 s and a room volume of about 85 m^3. The listening level was adjusted so that the highest reading of sound pressure level was 86 dBA (long integration time) at the centre of the front of the listening area.

The eight programme items listed in Table 1 were used to test all five codecs, resulting in a total of forty test presentations each for stereo and mono. The order of presentation of programme items and codecs was determined by a computer-generated random sequence, constrained to prevent the same programme item occurring in two consecutive tests and to ensure that all codecs were tested with all programme items. Two random sequences were generated, and used on alternate days.

The 'triple stimulus, hidden reference' method[9] was used for these tests. Listeners were presented with each test item in the sequence ABCABC, where A was always the unprocessed source material, or reference, and B and C comprised the processed material and another presentation of the reference; the listeners were not told which was which. An illuminated panel indicated to the listeners (17 in total) which of A, B or C was being presented.

Whether the hidden reference was in fact B or C was determined by another computer-generated random sequence, with the constraint that the number of tests in which it was B should equal the number of tests in which it was C, within two consecutive sessions each comprising ten tests. Listeners were asked to grade both B and C according to the CCIR 5-point impairment scale**[10]. The criterion for grading was specified as the perceived impairment of signal quality; no specific mention was made of possible distortion of the stereo image, although when the question was raised listeners were advised that they could make comments referring to the effect if it were noticed, and that it could influence their grading of an item.

Each morning or afternoon session was preceded by a recorded introduction explaining the method of presentation and the grading to be used. Two examples of significantly impaired material were demonstrated, together with the corresponding unprocessed source signals.

6. DISCUSSION OF RESULTS

At the conclusion of the listening tests, the mean of the grades awarded by the listeners was calculated for each codec processing each programme item, together with the standard deviation. The same calculations were performed for the hidden reference signals. The results are presented in the Appendix. Figs. A.1 to A.16 show the mean grades for both the processed signals and the hidden reference. Each diagram shows the results for all five codecs for

---

* A.K. McParland and K.J. Taylor assisted with the work described in this Report.

** The CCIR 5-point impairment scale is as follows:

5 Imperceptible
4 Perceptible, but not annoying
3 Slightly annoying
2 Annoying
1 Very annoying

a single programme item; Figs. A.1 to A.8 show the results for mono codecs, and Figs. A.9 to A.16 the results for stereo codecs. The results are summarised in Figs. A.17 to A.19, in which the mean of all the results for the different programme items are presented. Fig. A.17 shows the mean grades for the mono codecs, and Fig. A.18 the mean grades for the stereo codecs; the mean grades awarded to the reference signal are included for comparison. A comparison of the performance of the codecs in stereo and mono modes of operation is given in Fig. A.19, which incorporates the two previous sets of results (but without the results for the hidden reference).

In all tests, the mean grade for the hidden reference was consistently higher than 4.5, and was usually 4.7 or above. This is reassuring. Very occasionally, however, a grade below 4 was awarded to the hidden reference by one of the listeners.

No comments relating to distortion of the stereo image were received.

The codecs have been designated A, B, C, D and E in the results. Codecs D and E were, in fact, the same codec, operating at a sampling frequency of 32 kHz as codec D, and at 48 kHz as codec E. There were significant differences between the results obtained by codecs D and E with the same programme items, even though the overall results are similar. In the mono tests, codecs D and E performed consistently better than the others, with mean grades of 3.6. A significant improvement was evident in the performance of codec C in the stereo tests, relative to its performance in the mono tests, although it still did not reach the level of performance achieved by codecs D and E (overall mean grades of 3.95 and 3.7, respectively). The improvement was attributed to the success of codec C in exploiting stereo inter-channel correlation, in its 'joint stereo' mode of operation.

All codecs, except for codec A, had a 'joint stereo' mode. In the stereo tests, codec A operated in exactly the same mode as in the mono tests, but carried the two components of the stereo signal in two independent channels. It was awarded slightly lower grades in the stereo tests, however. It could be inferred from this that the listeners were judging codecs slightly more critically in the stereo tests. This, in turn, implies that all the other codecs were probably performing rather better in the stereo mode, relative to the mono mode, than is implied in the results of Figs. A.18 and A.19. Codec C showed the greatest improvement in going from the mono to the stereo mode.

Even the best of the codecs had difficulties with some types of signal; the lowest mean grade for codec D was 2.35, in the mono tests with castanets, and that for codec E was 2.0, in the stereo tests with harpsichord.

7. CONCLUSIONS

Subjective tests have been carried out on five very low bit-rate digital audio codecs. Although intended primarily for commentary applications, the codecs were evaluated using musical material as well as speech. All of the codecs introduced significant impairment on the most critical material, but two (codecs D and E) performed consistently better than the other three. However, even these were awarded particularly poor grades for the harpsichord and castanets programme items.

Codec C showed the greatest improvement in performance when going from the mono to the stereo mode of operation. It seems to be possible to make a significant improvement in the basic audio quality by adopting joint stereo coding techniques.

8. ACKNOWLEDGEMENTS

The authors are indebted to L. Mossberg of the Swedish Broadcasting Corporation and to J-Y. Leseure of the CCETT for their valuable participation in the task of selecting critical programme material.

Codecs were provided by members of the Eureka 147 partnership (the Institut für Rundfunktechnik (IRT), the CCETT, the Fraunhofer Gesellschaft, and Philips) and by Aware Inc.; the contribution made by these organisations to the work described in this Report is gratefully acknowledged.

9. REFERENCES

1. CCIR, 1990. Decision 98. Low bit-rate coding systems. Seventeenth Plenary Assembly (Düsseldorf, 1990) Green Book, Annex to **Vol. X**, Part 1, pp. 298-300.

2. CCIR, 1990. Recommendation 647. A digital audio interface for broadcasting studios. Seventeenth Plenary Assembly (Düsseldorf, 1990) Green Book, **Vol. XI**, Part 1, pp. 182-195.

3. GILCHRIST, N.H.C., 1990. Digital sound: subjective tests on low bit-rate codecs. BBC Research Department Report No. BBC RD 1990/16.

4. BERGMAN, S., GREWIN, C. *and* RYDÈN, T., 1990. MPEG/Audio subjective assessments test report. Swedish Broadcasting Corporation (SR) Research and Development Department Report, dated 30th July 1990, issued as CCIR Document JIWP 10-CMTT/1-20 (Study Period 1990-1994) and as ISO/IECJTC1/SC2/WG11 Document MPEG 90/196.

5. BERGMAN, S., GREWIN, C. *and* RYDÈN, T., 1991. The SR report on the MPEG/Audio subjective listening test, Stockholm, April/May 1991. Swedish Broadcasting Corporation (SR) Research and Development Department Report, dated 27th May 1991, issued as CCIR Document JIWP 10-CMTT/1-45 (Study Period 1990-1994).

6. KIRBY, D.G. *and* GENTLES, A.B., 1990. Digital audio editing: the development of a random access editor. BBC Research Department Report No. BBC RD 1990/5.

7. MASON, A.J. A digital audio matrix for HDTV. *BBC Research Department Report in preparation.*

8. GILCHRIST, N.H.C., 1993. Digital sound: the selection of critical programme material and preparation of recordings for CCIR tests on low bit-rate codecs. BBC Research Department Report No. BBC RD 1993/1.

9. GREWIN, C., BERGMAN, S. *and* KEJVING, O., 1986. A listening test system for evaluation of audio equipment. AES 80th Convention (Montreux, 1986). AES Preprint No. 2335.

10. CCIR, 1990. Recommendation 562. Subjective assessment of sound quality. Seventeenth Plenary Assembly (Düsseldorf, 1990) Green Book, **Vol. X**, Part 1, pp. 149-154.

APPENDIX

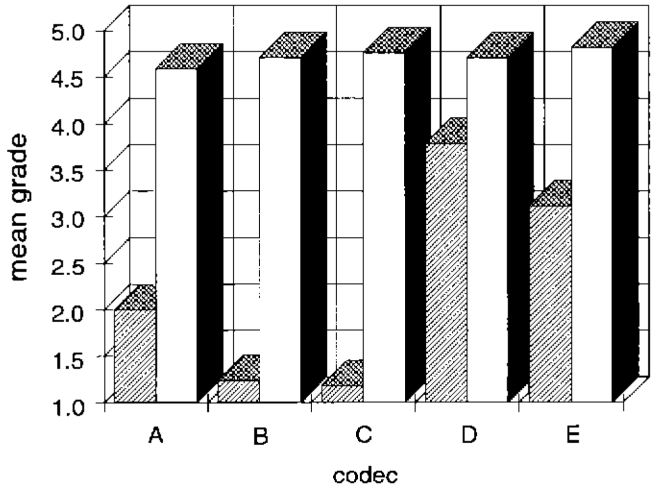


Fig. A.1 - Mean grades obtained with item 1 (Åsa Jinder): mono codecs.

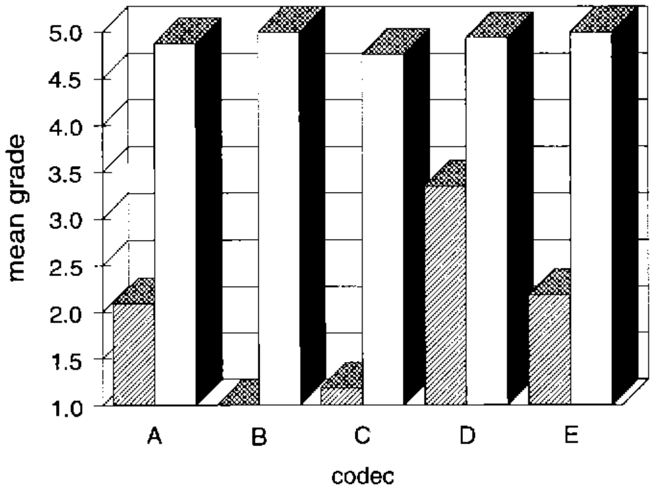


Fig. A.2 - Mean grades obtained with item 2 (harpsichord): mono codecs.

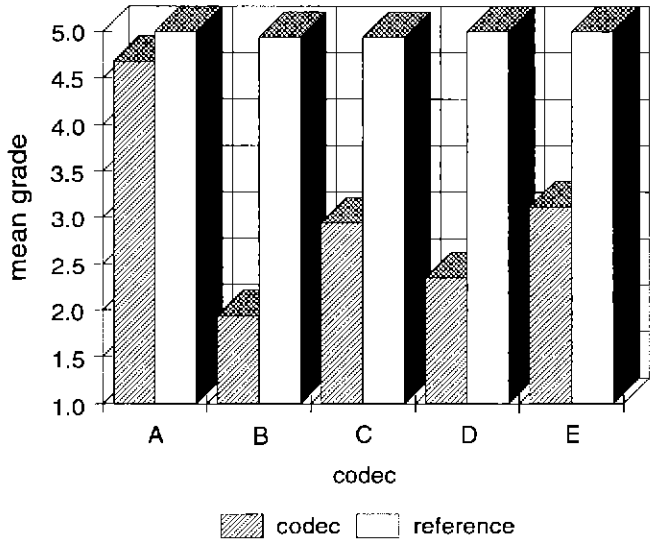


Fig. A.3 - Mean grades obtained with item 3 (castanets): mono codecs.

Fig. A.4 - Mean grades obtained with item 4 (male speech): mono codecs.

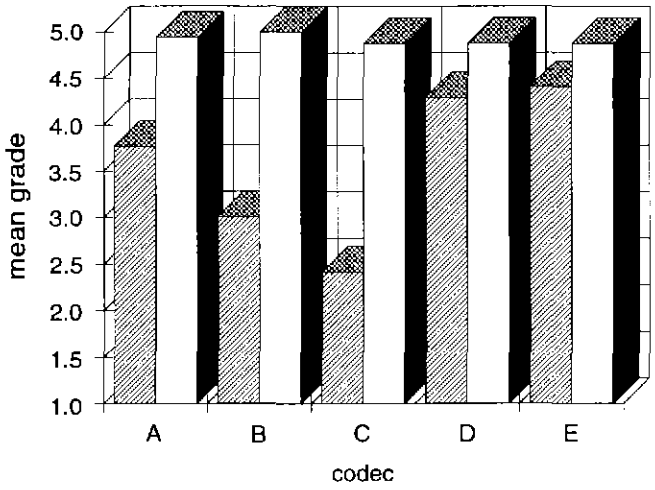


Fig. A.5 - Mean grades obtained with item 5 (bass clarinet): mono codecs.

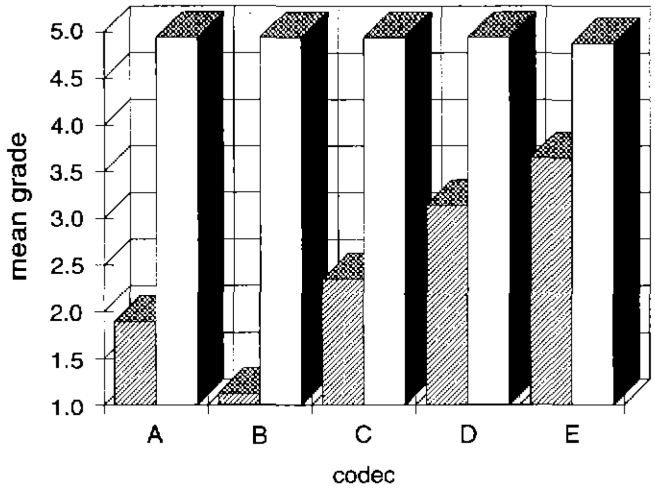
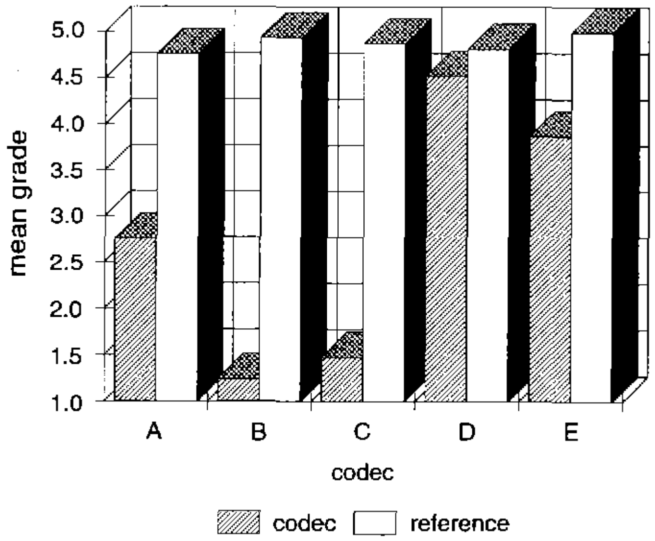


Fig. A.6 - Mean grades obtained with item 6 (Ornette Coleman): mono codecs.



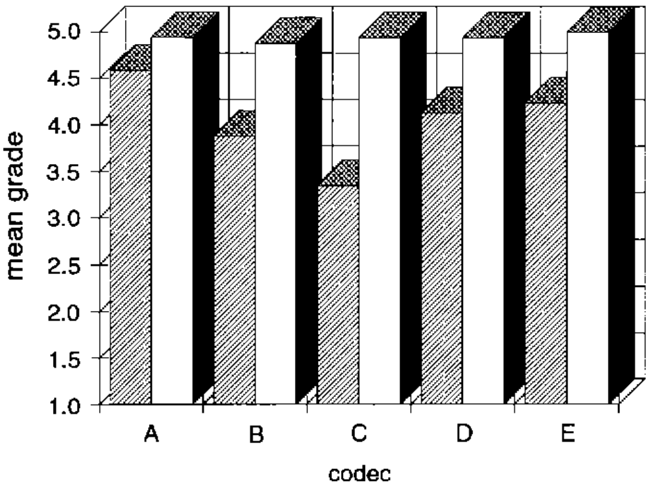


Fig. A.7 - Mean grades obtained with item 7 (bass guitar): mono codecs.

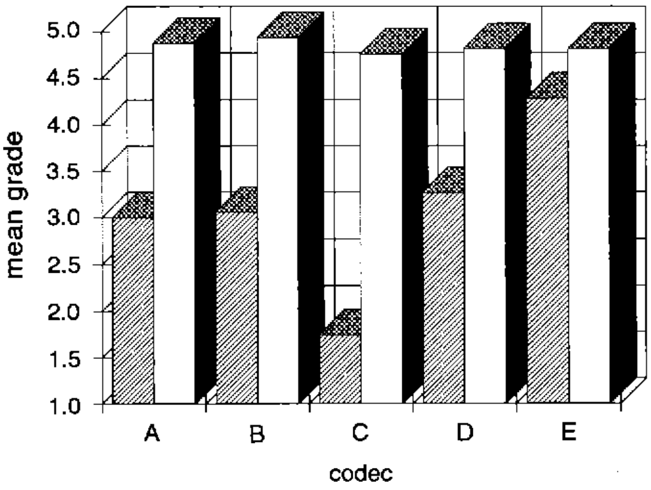


Fig. A.8 - Mean grades obtained with item 8 (Suzanne Vega): mono codecs.

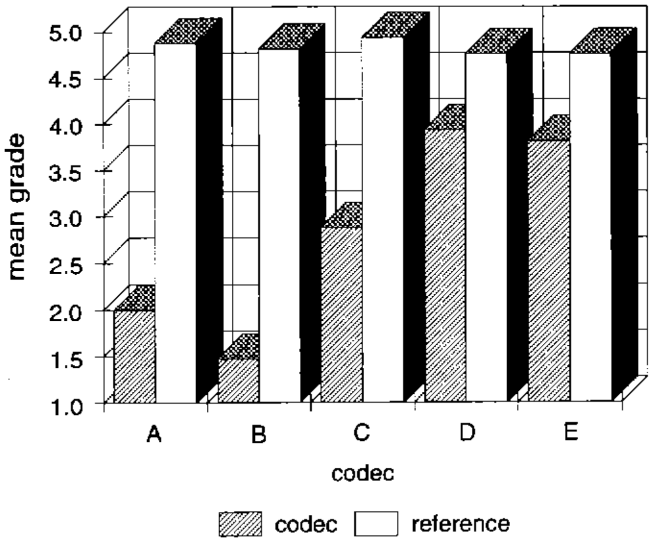


Fig. A.9 - Mean grades obtained with item 1 (Åsa Jinder): stereo codecs.

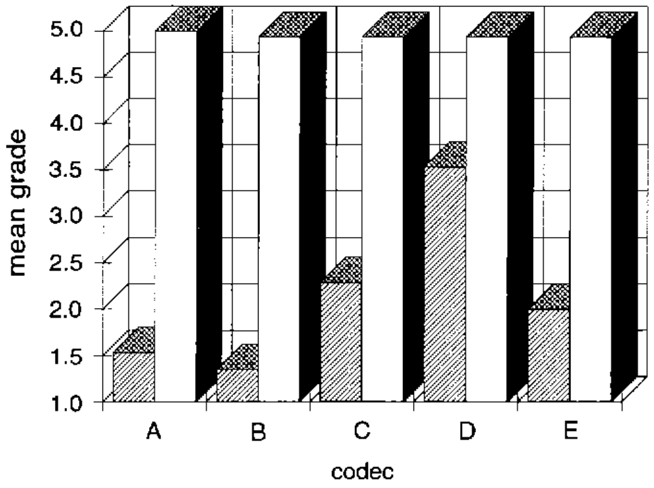


Fig. A.10 - Mean grades obtained with item 2 (harpsichord): stereo codecs.

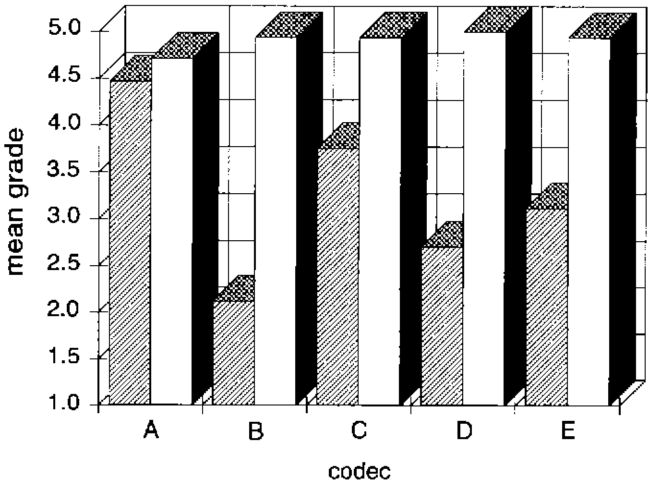


Fig. A.11 - Mean grades obtained with item 3 (castanets): stereo codecs.

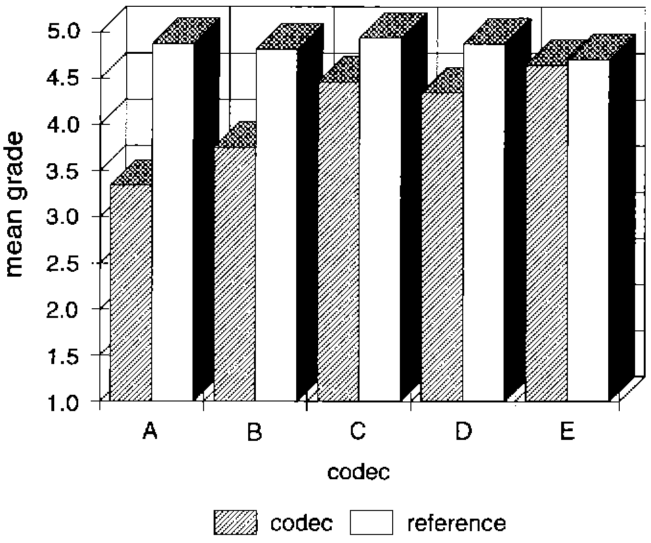


Fig. A.12 - Mean grades obtained with item 4 (male speech): stereo codecs.

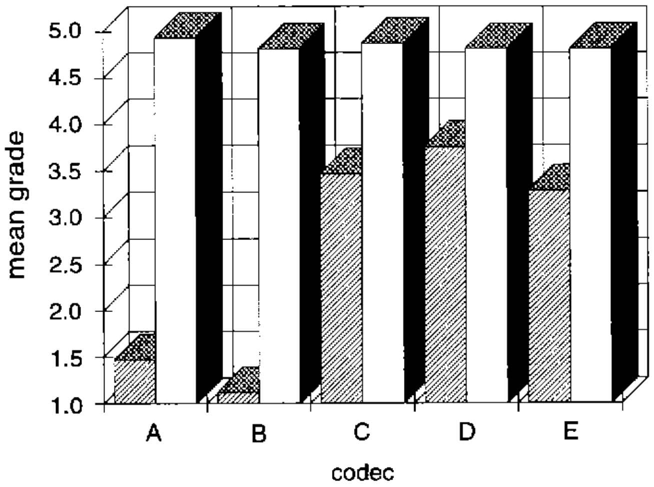


Fig. A.13 - Mean grades obtained with item 5 (bass clarinet): stereo codecs.

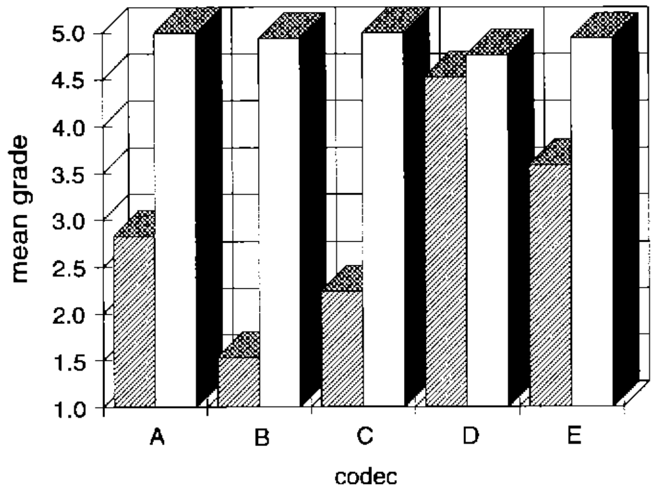


Fig. A.14 - Mean grades obtained with item 6 (Ornette Coleman): stereo codecs.

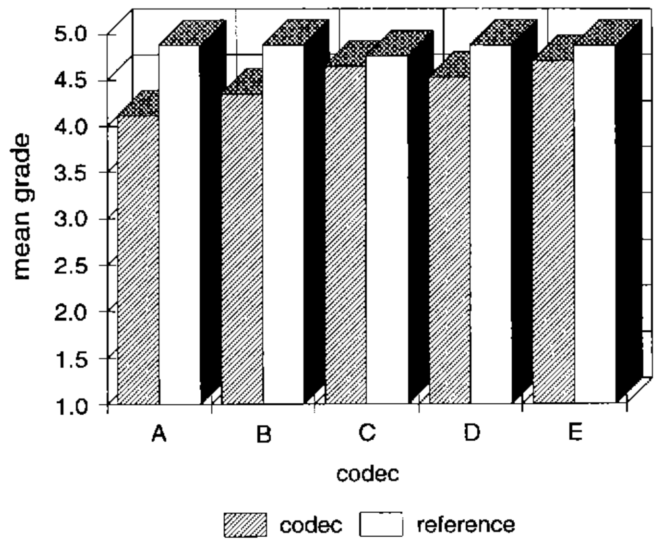


Fig. A.15 - Mean grades obtained with item 7 (bass guitar): stereo codecs.

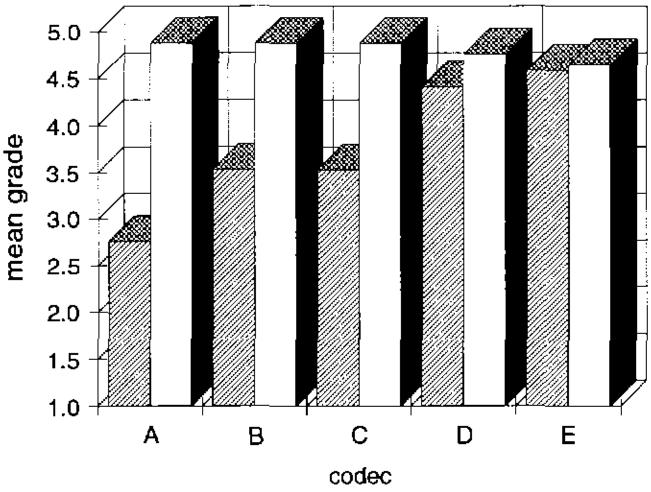


Fig. A.16 - Mean grades obtained with item 8 (Suzanne Vega): stereo codecs.

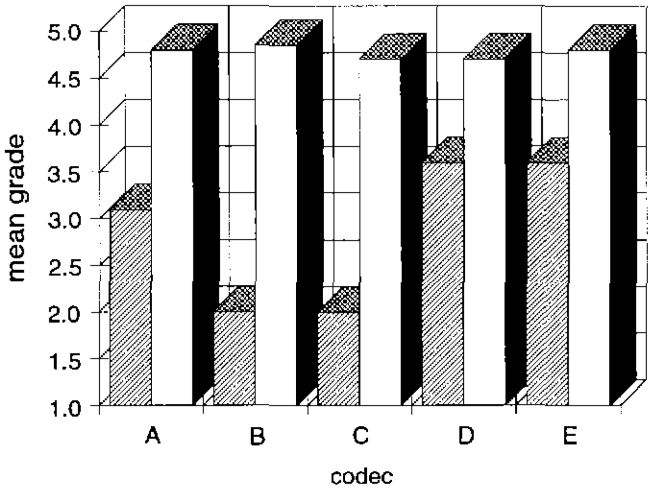


Fig. A.17 - Overall mean grades for mono codecs.

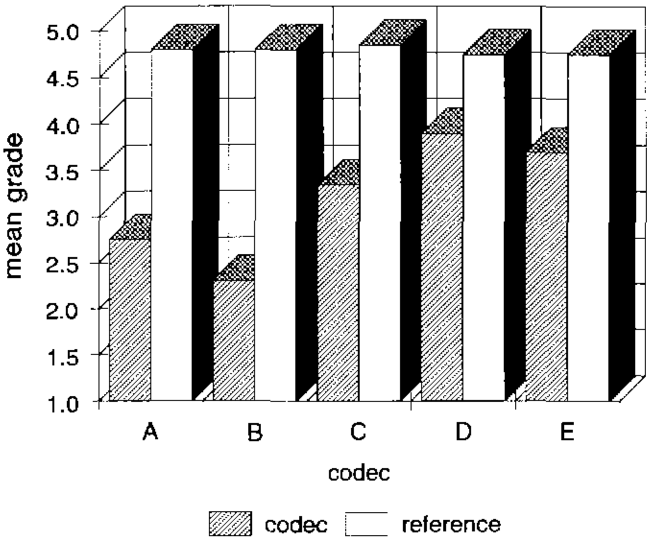


Fig. A.18 - Overall mean grades for stereo codecs.

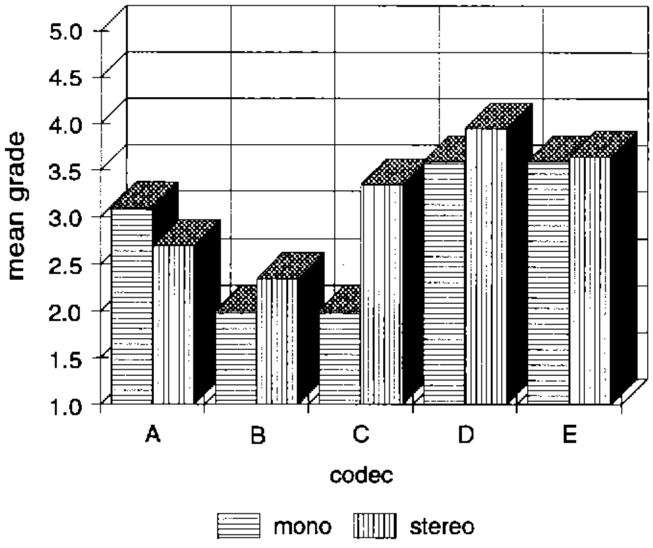


Fig. A.19 - Comparison of overall performance of mono and stereo codecs.

Printed by BBC RESEARCH DEPARTMENT, Kingswood Warren, Tadworth, Surrey, KT20 6NP